Theoretical Computer Science 81 (1991) 127-135
Elsevier

Note

An upward measure separation theorem *

Jack H. Lutz

Department of Computer Science, Iowa State University, Ames, IA 50011, USA

Communicated by J. Diaz
Received September 1989
Revised April 1990

Abstract

Lutz, J.H., An upward measure separation theorem (Note), Theoretical Computer Science 81 (1991) 127-135.

It is shown that almost every language in ESPACE is very hard to approximate with circuits. It follows that P ≠ BPP implies that E is a measure 0 subset of ESPACE.

1. Introduction

Hartmanis and Yesha [13] proved that P is a proper subset of P/Poly ∩ PSPACE if and only if E is a proper subset of ESPACE. (See Section 2 for notation and terminology used in this introduction.) This refined the *downward separation* result

$$E \subsetneq ESPACE \Rightarrow P \subsetneq PSPACE$$

of Book [4] and also led immediately to the *upward separation* result

$$P \subsetneq BPP \Rightarrow E \subsetneq ESPACE \tag{1.1}$$

of Hartmanis and Yesha [13]. (Work of Gill [9], Adleman [1], and Bennett and Gill [3] had already established that BPP is contained in P/Poly ∩ PSPACE.)

It is reasonable to conjecture that BPP is in fact a *proper* subset of P/Poly ∩ PSPACE, and hence that the $P \subsetneq BPP$ hypothesis might yield a stronger conclusion

* This work was supported in part by NSF Grant CCR-8809238.

than the separation of E from ESPACE. This paper supports this intuition by proving the following.

Main Theorem. *If* $P \subsetneq BPP$, *then* $\mu(E|ESPACE) = 0$.

The conclusion here states that E is a *measure 0*, i.e. negligibly small, subset of ESPACE in the resource-bounded measure theory of Lutz [20, 21]. (This theory, which has the *classical* and *effective* Lebesgue measure theories (cf. [10, 8, 23]) as special cases, describes the internal measure-theoretic structure of ESPACE and other complexity classes.) Thus the Main Theorem is an *upward measure separation* result which extends (1.1) by asserting that *any* separation of P from BPP implies a *measure separation* of E from ESPACE.

The proof of the Main Theorem makes essential use of two recent results, presented as Theorems 1 and 2 below. Theorem 1, from Nisan and Wigderson [24, 25], states that P = BPP if E contains any problem "with hardness $2^{\alpha n}$ for some $\alpha > 0$". Theorem 2, from Lutz [20], states that almost every problem in ESPACE has "high selective space-bounded Kolmogorov complexity" almost everywhere. Precise statements of these theorems, together with necessary definitions, are given in Section 3. The proofs of Theorems 1 and 2, which involve pseudorandom bit generators and resource-bounded measure theory, respectively, are not repeated here. In fact, Theorem 2 captures all the resource-bounded measure theory needed for the Main Theorem, so no measure theory is used in this paper. Details of resource-bounded measure theory may be found in [20, 21] but such details are not needed to follow the argument of this paper.

In Section 4, Theorem 2 is used to prove Theorem 3, which states that almost every problem in ESPACE "has hardness greater than $2^{\alpha n}$ for every $0 < \alpha < \frac{1}{3}$", i.e., is very hard to approximate with circuits. The Main Theorem follows immediately from Theorems 1 and 3.

2. Preliminaries

All results in this paper are robust with respect to reasonable choices of the underlying model of computation. Our *machines* can thus be interpreted as Turing machines, pointer machines, random access machines, etc.

All *languages* here are sets $L \subseteq \{0, 1\}^*$. We write $L_{=n}$ for $L \cap \{0, 1\}^n$. The *characteristic string* of $L_{=n}$ is the 2^n-bit string $\chi_{L_{=n}}$ whose ith bit is 1 iff $w_i \in L$, where w_i is the ith string in the lexicographic enumeration of $\{0, 1\}^n$. We write $|x|$ for the *length* of a string $x \in \{0, 1\}^*$.

The *symmetric difference* of sets A and B is denoted by $A \triangle B = (A \setminus B) \cup (B \setminus A)$. The *cardinality* of a finite set A is denoted by $|A|$.

Our *circuits* are Boolean, combinational (acyclic) circuits with bounded fan-in, unbounded fan-out, and a single output gate. An n-input circuit γ *computes* the set

$L(\gamma)$ of all strings $w \in \{0,1\}^n$ for which $\gamma(w)$, the Boolean value of the output gate on input w, is 1. The *size* of a circuit γ, written size(γ), is the number of gates in γ. The *circuit-size complexity* of a language L is the function $\mathrm{CS}_L : \mathbf{N} \to \mathbf{N}$ defined by

$$\mathrm{CS}_L(n) = \min\{\mathrm{size}(\gamma) \mid L(\gamma) = L_{=n}\}.$$

Further details (which are standard and can be varied in minor ways) may be found in [2, 20] or any standard reference on circuit complexity.

We are interested in the polynomial complexity classes P and PSPACE, the exponential complexity classes $\mathrm{E} = \mathrm{DTIME}(2^{\mathrm{linear}})$ and $\mathrm{ESPACE} = \mathrm{DSPACE}(2^{\mathrm{linear}})$, the bounded-error probabilistic time complexity class BPP defined by Gill [9], and the nonuniform complexity class

$$\mathrm{P/Poly} = \{L \mid \mathrm{CS}_L(n) = n^{\mathrm{O}(1)}\},$$

consisting of all languages which have polynomial-size circuits.

A property $\varphi(n)$ of natural numbers n holds *infinitely often* (i.o.) if it holds for infinitely many $n \in \mathbf{N}$, and *almost everywhere* (a.e.) if it holds for all but finitely many $n \in \mathbf{N}$.

In Section 4 we use (a special case of) the Chernoff bound [6] which can be found in [7, 19] and many other references. This result states that

$$\sum_{0 \le i \le aN} \binom{N}{i} p^i (1-p)^{N-i} \le \rho^N \tag{2.1}$$

for all $0 < a < p < 1$, where

$$\rho = \left(\frac{p}{a}\right)^a \left(\frac{1-p}{1-a}\right)^{1-a}.$$

If we set $p = \frac{1}{2}$, then (2.1) tells us that

$$\sum_{0 \le i \le aN} \binom{N}{i} \le 2^N \rho^N. \tag{2.2}$$

We will use (2.2) in the case where $p = \frac{1}{2}$ and $a = \frac{1}{2}(1-\varepsilon)$ for some $\varepsilon > 0$. In this case,

$$\rho = [(1-\varepsilon)^{\varepsilon-1}(1+\varepsilon)^{-\varepsilon-1}]^{1/2} = \left[(1-\varepsilon^2)^{-1}\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{\varepsilon}\right]^{1/2}.$$

3. Two recent results

This section summarizes two recent results which are used to prove the upward measure separation.

Definition (Nisan and Wigderson [24, 25]). Given $\delta > 0$ and $n, s \in \mathbf{N}$, a language $L \subseteq \{0,1\}^*$ is (δ, s)-*hard at* n if

$$|L(\gamma) \triangle L_{=n}| > 2^{n-1}(1-\delta)$$

for every n-input circuit γ with size$(\gamma) \leq s$. The *hardness* of a language $L \subseteq \{0,1\}^*$ is the function $H_L : \mathbf{N} \to \mathbf{N}$ defined by

$$H_L(n) = \max\{h \in \mathbf{N} \mid L \text{ is } (h^{-1}, h)\text{-hard at } n\}.$$

Thus a language L is (δ, s)-hard at n if every n-input circuit of size s computes L incorrectly on at least $50(1-\delta)$ percent of the inputs in $\{0,1\}^n$. Note that $H_L(n)$ is bounded above by the size of the smallest circuit which correctly computes $L_{=n}$.

For each $0 < \alpha < 1$, we define the set

$$H_\alpha = \{L \subseteq \{0,1\}^* \mid H_L(n) > 2^{\alpha n} \text{ a.e.}\}$$

of languages with hardness greater than $2^{\alpha n}$ almost everywhere.

A new construction of a pseudorandom bit generator was recently used to prove the following theorem.

Theorem 1 (Nisan and Wigderson [24, 25]). *If $E \cap H_\alpha \neq \emptyset$ for some $\alpha > 0$, then* P = BPP.

The second result which we review in this section is (a special case of) an almost everywhere lower bound on the space-bounded Kolmogorov complexity of languages in ESPACE. (Kolmogorov complexity was originally introduced by Solomonoff [28], Kolmogorov [15], and Chaitin [5]. Time- and space-bounded Kolmogorov complexities have since been investigated by Hartmanis [11], Sipser [27], Levin [17], Huynh [12], Ko [14], Longpré [18], Lutz [20, 21], and many others. For an overview of work in this area, see Kolmogorov and Uspenskii [16] or Li and Vitányi [22].)

Definition. Given a machine M, a resource bound $t : \mathbf{N} \to \mathbf{N}$, a language $L \subseteq \{0,1\}^*$, and a natural number n, the *t-space-bounded Kolmogorov complexity* of $L_{=n}$ relative to M is

$$KS^t_M(L_{=n}) = \min\{|\pi| \mid M(\pi, n) = \chi_{L_{=n}} \text{ in } \leq t(2^n) \text{ space}\},$$

i.e., the length of the shortest program π such that M, on input (π, n), outputs the characteristic string of $L_{=n}$ and halts without using more than $t(2^n)$ workspace.

Well-known simulation techniques show that there exists a machine U which is *optimal* in the sense that for each machine M there is a constant c such that for all t, L, and n we have

$$KS^{ct+c}_U(L_{=n}) \leq KS^t_M(L_{=n}) + c.$$

As usual, we fix an optimal machine U and omit it from the notation.

It can easily be seen that if $x \in \{0,1\}^\infty$ is the characteristic sequence of $L \subseteq \{0,1\}^*$, then $KS^t(L_{=n})$ is precisely $KS^t(x \wedge \hat{\sigma} \mid 2^{n+1} - 1)$, the t-space bounded $\hat{\sigma}$-selective Kolmogorov complexity of x, as defined in [20]. We thus have the following result.

Theorem 2 (Lutz [20]). *For any polynomial q and any real $a > 1$, if*

$$X = \{L \subseteq \{0,1\}^* \mid \mathrm{KS}^q(L_{=n}) > 2^n - an \text{ a.e.}\},$$

then $\mu(X \mid \mathrm{ESPACE}) = 1$.

The conclusion of Theorem 2 says that *almost every* language in ESPACE is in X, i.e., has high q-space bounded Kolmogorov complexity almost everywhere. A precise definition of the condition $\mu(X \mid \mathrm{ESPACE}) = 1$ may be found in [20, 21], but is not needed here because Theorem 2 gives us the means to prove a variety of measure-theoretic results without explicitly discussing measure.

The only other properties of measure which we use are the following trivial facts.

(i) If $X \subseteq Y$ and $\mu(X \mid \mathrm{ESPACE}) = 1$, then $\mu(Y \mid \mathrm{ESPACE}) = 1$.

(ii) If $X \cap Y = \emptyset$ and $\mu(X \mid \mathrm{ESPACE}) = 1$, then $\mu(Y \mid \mathrm{ESPACE}) = 0$.

Beyond this, we hope that the reader will accept (or acquire from [20, 21]) the intuition that $\mu(X \mid \mathrm{ESPACE}) = 0$ means that $X \cap \mathrm{ESPACE}$ is a very small subset of ESPACE.

4. Upward measure separation

The following result is the technical content of this section.

Theorem 3. *If* $\mathrm{H} = \bigcap_{0<\alpha<1/3} \mathrm{H}_\alpha$, *then* $\mu(\mathrm{H} \mid \mathrm{ESPACE}) = 1$.

This result is interesting in and of itself, since it says that almost every language in ESPACE is very hard to approximate with circuits. In this paper we are especially interested in the following application.

Main Theorem. *If* $\mathrm{P} \subsetneq \mathrm{BPP}$, *then* $\mu(\mathrm{E} \mid \mathrm{ESPACE}) = 0$.

Proof. Let H be as in Theorem 3. If $\mathrm{P} \subsetneq \mathrm{BPP}$, then $\mathrm{E} \cap \mathrm{H} = \emptyset$ by Theorem 1. Since $\mu(\mathrm{H} \mid \mathrm{ESPACE}) = 1$, it follows that $\mu(\mathrm{E} \mid \mathrm{ESPACE}) = 0$. □

Thus any separation of P from BPP implies a measure separation of E from ESPACE.

The rest of this section is devoted to the proof of Theorem 3. We use the following lemmas.

Lemma 4. *For any real $b < 1$, for all sufficiently small reals $\varepsilon > 0$,*

$$(1-\varepsilon^2)^{-1}\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{\varepsilon} < 1 - b\varepsilon^2.$$

Lemma 5. *There exist a polynomial q and a constant $c>0$ with the following property. For every two reals $0<\alpha<\beta<1$, for all sufficiently large n, for every language $L\subseteq\{0,1\}^*$, if $H_L(n)\leq 2^{\alpha n}$, then*

$$KS^q(L_{=n})<2^n-c2^{(1-2\alpha)n}+2^{\beta n}.$$

Proof of Theorem 3. Choose q and c as in Lemma 5 and define X as in Theorem 2, using $a=2$. We will show that $X\subseteq H$, whence Theorem 3 follows from Theorem 2.

Assume that $L\notin H$, i.e., that $L\notin H_\alpha$ for some $0<\alpha<\frac{1}{3}$. Fix β such that $\alpha<\beta<1-2\alpha$. Then $H_L(n)\leq 2^{\alpha n}$ i.o., so the inequality in the conclusion of Lemma 5 holds for infinitely many n. Since $\beta<1-2\alpha$, the right-hand side of this inequality is less than 2^n-2n for all sufficiently large n, so it follows that $L\notin X$. □

Proof of Lemma 4. Calculating with Taylor approximations, we have

$$\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^\varepsilon=(1-2\varepsilon+o(\varepsilon))^\varepsilon=e^{\varepsilon\ln(1-2\varepsilon+o(\varepsilon))}$$

$$=e^{-2\varepsilon^2+o(\varepsilon^2)}=1-2\varepsilon^2+o(\varepsilon^2)$$

as $\varepsilon\to 0$. Since $b<1$ and $(1-\varepsilon^2)(1-b\varepsilon^2)=1-(1+b)\varepsilon^2+o(\varepsilon^2)$ as $\varepsilon\to 0$, it follows that

$$\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^\varepsilon<(1-\varepsilon^2)(1-b\varepsilon^2)$$

for all sufficiently small ε. □

Proof of Lemma 5. Call an n-input circuit γ *novel* if no n-input circuit which precedes γ (in a standard enumeration of all circuits; no circuit precedes a smaller one in this enumeration) computes the same set as γ. The predicate "γ is a novel n-input circuit" can clearly be tested in space which is polynomial in $n+\text{size}(\gamma)$. Let $\gamma_1,\ldots,\gamma_{J(n)}$ be the enumeration of all novel n-input circuits (in their order of appearance in the standard enumeration). Also, let $N=2^n$ and let $\Delta_1,\ldots,\Delta_{J(n)}$ be the enumeration of $\{0,1\}^N$ which is lexicographic, except that no string precedes a string which has fewer 1s. (Of course, $J(n)=2^N=2^{2^n}$ in both cases.) It is routine to design a machine M which takes inputs $\pi\in\{0,1\}^*$ and $n\in\mathbf{N}$ and has the following property. If $\pi=\langle t,d\rangle$, where $t,d\in\{1,\ldots,J(n)\}$ are represented in binary, then $M(\pi,n)=\text{graph}(\gamma_t)\oplus\Delta_d$, where $\text{graph}(\gamma_t)$ is the N-bit characteristic string of the set computed by γ_t, $\oplus$ denotes bitwise exclusive-or, and this computation is carried out in space which is polynomial in 2^n. Since the pairing function can be implemented with $|\langle t,d\rangle|\leq|t|+|d|+2\log|t|+4$, and since we have fixed an optimal machine in defining KS, it follows that there exist a polynomial q and a constant c_1 such that

$$KS^q(L_{=n})\leq|t|+|d|+2\log|t|+c_1 \tag{4.1}$$

whenever $\text{graph}(\gamma_t)\oplus\Delta_d$ is the characteristic string of $L_{=n}$.

Now fix $0<\alpha<\beta<1$. A standard counting argument (see, for example [26, 2, 20]) shows that at most $[48e2^{\alpha n}]^{2^{\alpha n}}=[48eN^{\alpha}]^{N^{\alpha}}$ n-input circuits γ are novel and have size$(\gamma)\leq 2^{\alpha n}$. The number $D(n)$ of N-bit strings Δ which have $\frac{1}{2}N(1-N^{-\alpha})=2^{n-1}(1-2^{-\alpha n})$ or fewer 1s is given by

$$D(n)=\sum_{0\leq i\leq aN}\binom{N}{i},$$

where we write $a=\frac{1}{2}(1-\varepsilon)$ and $\varepsilon=N^{-\alpha}$ for convenience. By the Chernoff bound [6] discussed in Section 2, this implies that

$$D(n)\leq 2^N\rho^N,$$

where

$$\rho=\left[(1-\varepsilon^2)^{-1}\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{\varepsilon}\right]^{1/2}.$$

It follows by Lemma 4 that

$$D(n)\leq 2^N(1-\tfrac{1}{2}\varepsilon^2)^{N/2}=2^{N+(N/2)\log(1-\varepsilon^2/2)}$$

for all sufficiently large n. Since

$$\log(1-\tfrac{1}{2}\varepsilon^2)=\frac{1}{\ln 2}\ln(1-\tfrac{1}{2}\varepsilon^2)\leq\frac{-\varepsilon^2}{2\ln 2}$$

for all ε, it follows that

$$D(n)\leq 2^{N-cN\varepsilon^2}=2^{N-cN^{1-2\alpha}} \tag{4.2}$$

for all sufficiently large n, where $c=1/4\ln 2$.

Now let n be large enough that (4.2) holds and

$$2+\log K+2\log(1+\log K)+c_1<N^{\beta}, \tag{4.3}$$

where $K=[48eN^{\alpha}]^{N^{\alpha}}$ and c_1 is as in (4.1). Assume that $\mathrm{H}_L(n)\leq 2^{\alpha n}$. Then, by (4.2) and our estimate of the number of novel circuits of size $\leq 2^{\alpha n}$, there exist $t\leq K$ and $d\leq 2^{N-cN^{1-2\alpha}}$ such that graph $(\gamma_t)\oplus\Delta_d$ is the characteristic string of $L_{=n}$. It follows by (4.1) and (4.3) that

$$\begin{aligned}\mathrm{KS}^q(L_{=n})&\leq|t|+|d|+2\log|t|+c_1\\&\leq 1+\log K+1+N-cN^{1-2\alpha}+2\log(1+\log K)+c_1\\&<N-cN^{1-2\alpha}+N^{\beta}\\&=2^n-c2^{(1-2\alpha)n}+2^{\beta n}.\end{aligned}$$

□

5. Conclusion

This paper refines the picture

$$\mathrm{P}\subsetneq\mathrm{BPP}\Rightarrow\mathrm{P}\subsetneq\mathrm{P/Poly}\cap\mathrm{PSPACE}\Leftrightarrow\mathrm{E}\subsetneq\mathrm{ESPACE}$$

to the form

$$\begin{array}{ccc} P \subsetneq BPP & \Rightarrow & \mu(E \mid ESPACE) = 0 \\ \Downarrow & & \Downarrow \\ P \subsetneq P/Poly \cap PSPACE & \Leftrightarrow & E \subsetneq ESPACE. \end{array}$$

It will be interesting to see the situation clarified further.

Acknowledgment

I thank Noam Nisan, Giora Slutzki, William Schmidt, and David Juedes for helpful discussions.

References

[1] L. Adleman, Two theorems on random polynomial time, in: *Proc. 19th IEEE Symp. on Foundations of Computer Science* (1978) 75-83.
[2] J.L. Balcázar, J. Díaz, and J. Gabarró, *Structural Complexity I* (Springer, Berlin, 1988).
[3] C.H. Bennett and J. Gill, Relative to a random oracle A, $P^A \neq NP^A \neq co\text{-}NP^A$ with probability 1, *SIAM J. Comput.* **10** (1981) 96-113.
[4] R.V. Book, Comparing complexity classes, *J. Comput. System Sci.* **9** (1974) 213-229.
[5] G.J. Chaitin, On the length of programs for computing finite binary sequences, *J. Assoc. Comput. Mach.* **13** (1966) 547-569.
[6] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.* **23** (1952) 493-507.
[7] P. Erdös and J. Spencer, *Probabilistic Methods in Combinatorics* (Academic Press, New York, 1974).
[8] R.I. Freidzon, Families of recursive predicates of measure zero, translated in: *J. Soviet Math.* **6** (1976) 449-455.
[9] J. Gill, Computational complexity of probabilistic Turing machines, *SIAM J. Comput.* **6** (1977) 675-695.
[10] P.R. Halmos, *Measure Theory* (Springer, Berlin, 1950).
[11] J. Hartmanis, Generalized Kolmogorov complexity and the structure of feasible computations, in: *Proc. 24th IEEE Symp. on the Foundations of Computer Science* (1983) 439-445.
[12] D.T. Huynh, Resource-bounded Kolmogorov complexity of hard languages, in: *Structure in Complexity Theory* (Springer, Berlin, 1986) 184-195.
[13] J. Hartmanis and Y. Yesha, Computation times of NP sets of different densities, *Theoret. Comput. Sci.* **34** (1984) 17-32.
[14] K. Ko, On the notion of infinite pseudorandom sequences, *Theoret. Comput. Sci.* **48** (1986) 9-33.
[15] A.N. Kolmogorov, Three approaches to the quantitative definition of "information", *Problems Inform. Transmission* **1** (1965) 1-7.
[16] A.N. Kolmogorov and V.A. Uspenskii, Algorithms and randomness, translated in: *Theory Probab. Appl.* **32** (1987) 389-412.
[17] L.A. Levin, Randomness conservation inequalities; information and independence in mathematical theories, *Inform. and Control* **61** (1984) 15-37.
[18] L. Longpré, Resource bounded Kolmogorov complexity, a link between computational complexity and information theory, Ph.D. Thesis, Technical Report TR-86-776, Cornell University, 1986.
[19] J.H. Lutz, An elementary combinatorial derivation of "the" Chernoff bound, Technical Report 88-8, Iowa State University, 1988.

[20] J.H. Lutz, Almost everywhere high nonuniform complexity, *J. Comput. System Sci.* to appear. Also in: *Proc. 4th Structure in Complexity Theory Conf.* (1989) 81-91.

[21] J.H. Lutz, Category and measure in complexity classes, *SIAM Journal on Computing* **19** (1990) 1100-1131.

[22] M. Li and P.M.B. Vitányi, Two decades of applied Kolmogorov complexity: In memoriam Andrei Nikolaevich Kolmogorov 1903-1987, in: *Proc. 3rd Structure in Complexity Theory Conf.* (1988) 80-101.

[23] K. Mehlhorn, The "almost all" theory of subrecursive degrees is decidable, in: *Proc. 2nd Coll. on Automata, Languages, and Programming* (1974) 317-325.

[24] N. Nisan and A. Wigderson, Hardness vs. randomness, in: *Proc. 29th IEEE Symp. on Foundations of Computer Science* (1988) 2-11.

[25] N. Nisan and A. Wigderson, Hardness vs. randomness, in preparation.

[26] C.E. Shannon, The synthesis of two-terminal switching circuits, *Bell System Tech. J.* **28** (1949) 59-98.

[27] M. Sipser, A complexity-theoretic approach to randomness, in: *Proc. 15th ACM Symp. on the Theory of Computing* (1983) 330-335.

[28] R.J. Solomonoff, A formal theory of inductive inference, *Inform. and Control* **7** (1964) 1-22; 224-254.